

# Raportti vuoden 1998 toiminnasta



Helsingin yliopiston elektroninen julkaisupankki

# Sisällysluettelo

1.	Yhteenveto vuoden 1998 toiminnasta .....	1
2.	Taustaselvitykset .....	2
2.1	Elektronisen julkaisemisen nykytila yliopistossa.....	2
2.1.1	Teologinen tiedekunta .....	2
2.1.2	Oikeustieteellinen tiedekunta .....	2
2.1.3	Humanistinen tiedekunta .....	2
2.1.4	Matemaattis-luonnontieteellinen tiedekunta .....	3
2.1.5	Kasvatustieteellinen tiedekunta .....	3
2.1.6	Valtiotieteellinen tiedekunta.....	3
2.2	Muita vastaavia koti- ja ulkomaisia hankkeita .....	4
2.2.1	HUTPubl .....	4
2.2.2	Oulun yliopiston kirjasto – elektroniset julkaisut.....	4
2.2.3	Jyväskylän yliopiston kirjaston tutkielmapankki .....	4
2.2.4	The Nordic Centre of Excellence for Electronic Publishing (NordEP) ...	4
2.2.5	Linköping University Electronic Press.....	5
2.2.6	The Internet Project .....	5
2.2.7	NYST-Informasjonsformidling .....	5
2.2.8	NDLTD — Networked Digital Library of Theses and Dissertations .....	6
2.2.9	UW Electronic Theses and Dissertations Server.....	6
2.3	Väitöskirjan julkaiseminen.....	6
3.	Sopimukset .....	7
4.	Julkaisu ympäristö .....	8
5.	Julkaisuohjeet ja julkaisu tuotanto .....	9
5.1	Pilottiaineisto.....	9
5.2	Yhtenäinen ulkoasu .....	9
5.3	Tuettavat ohjelmistot .....	9
5.4	DC-kuvailut ja URN-generaattori .....	10
5.5	Ohjeet kirjoittajille .....	10
5.6	Tiedekuntien toimittama aineisto .....	11
6.	Julkaisujen arkistointi.....	12
7.	Koulutus ja tiedotus.....	12
8.	Määrärahan käyttö .....	13

# 1. Yhteenveto vuoden 1998 toiminnasta

Helsingin elektroninen julkaisupankki -hanke käynnistyi vuoden 1998 alussa. Hankkeen johtoryhmä asetettiin 3.3.1998 ja projektisuunnitelma hyväksyttiin 25.5.1998. Johtoryhmä on vuoden aikana kokoontunut 4 kertaa. Johtoryhmä on perustanut erilliset työryhmät tekijänoikeussopimusten valmistelua ja väitöskirjojen elektronista julkaisemista varten. Molemmat työryhmät ovat kokoontuneet kaksi kertaa.

Hankkeen johtana on Inkeri Salonharju (HYK). Johtoryhmän puheenjohtajana toimii Antti Kivivuori (yksityisoikeuden laitos) ja sihteerinä Timo Hellgren (HYK). Muut johtoryhmän jäsenet ovat: Paul Fogelberg, Jukka Englund (terveys-tieteiden keskuskirjasto), Hannele Kauranne (yliopistopaino), Jarmo Koskela (yliopiston tiedotus), Kimmo Koskeniemi (yleisen kielitieteen laitos), Markku Leskelä (kemian laitos), Ari Muhonen (HYK), Heli Myllys (maatalouskirjasto), ja Martti Tammisto (atk-osasto).

Hankkeen tavoitteena on

- luoda julkaisu- ja välityspalvelu yliopiston omia elektronisia julkaisuja varten
- kehittää sopimusmekanismit ja mallisopimukset tekijänoikeuksien hankkimiseksi ja hallinnoimiseksi
- laatia elektronisen julkaisemisen edellyttämä ohjeistus julkaisujen kirjoittajille
- luoda kontaktit tiedekuntiin ja kehittää automaattiset kanavat aineiston siirtämiseksi julkaisupalveluun ja arkistoon
- perustaa elektronisen julkaisuaineiston arkisto
- integroida palveluun tarvepainatus.

Toimintavuoden aikana on tehty selvitykset Helsingin yliopiston elektronisesta julkaisutilanteesta, ja selvitetty koti- ja ulkomaisia vastaavia hankkeita, rakennettu elektronisen julkaisemisen ympäristö, laadittu sopimusluonnokset tekijänoikeussopimuksiksi, laadittu ohjeiden ensimmäiset versiot julkaisujen kirjoittajille aineiston luovuttamista, ulkoasua ja julkaisuun sisällytettäviä kuvailutietoja (URN, DC-kuvailu) varten, luotu arkistorakenne julkaisujen arkistointia varten, avattu hankkeen kotisivu, käynnistetty tiedekuntien yhdyshenkilöverkosto ja koulutettu yhteyshenkilöitä.

## 2. Taustaselvitykset

### 2.1 Elektronisen julkaisemisen nykytila yliopistossa

Helsingin yliopiston elektroninen julkaisupankki teki 27.4.1998 selvityksen yliopiston WWW-sivuilla olevista laitosten ja tiedekuntien toimittamista tieteellisistä elektronisista julkaisuista. Selvitykseen sisältyvät vain ne julkaisut, joista on linkki laitoksen tai tiedekunnan sivuilta. Opiskelijoiden tai henkilökunnan sivuilla olevaa materiaalia ei selvitetty. Tulokset ovat esillä myös verkossa, osoitteessa: <http://linnea.helsinki.fi/elbanco/publicat.html> Ne tiedekunnat, joita ei ole erikseen mainittu, eivät ole julkaisseet mitään elektronisessa muodossa.

#### 2.1.1 Teologinen tiedekunta

Tiedekunnan kolme laitosta (kirkkohistorian laitos, systemaattisen teologian laitos ja käytännöllisen teologian laitos) ovat julkaisseet verkossa ohjeita opiskelijoille HTML-muodossa.

#### 2.1.2 Oikeustieteellinen tiedekunta

Tiedekunnan sivuilla on Pekka Timosen *Johdatus lainopin metodiin ja lainopilliseen kirjoittamiseen* HTML- ja PDF-versioina (<http://www.helsinki.fi/oik/tdk/tutkopas/>)

#### 2.1.3 Humanistinen tiedekunta

Tiedekunnan laitoksista Aasian ja Afrikan kielten laitos (6 kpl), fonetiikan laitos (1 kpl) sekä kotimaisen kirjallisuuden laitos (2 kpl) ovat julkaisseet tieteellisiä artikkeleita HTML-muodossa. Kognitiotieteen laitos on julkaissut opiskelijoiden artikkeleita PDF-muodossa (17 kpl).

Fonetiikan laitoksen julkaisu on tehty alkuaan LaTeX:lla ja käännetty siitä HTML:ksi latex2html-ohjelmalla. Kotimaisen kirjallisuuden laitoksen julkaisema väitöskirja on muunnettu PageMakerilla HTML:ksi. Samasta työstä on PDF-versio elektronisen julkaisupankin palvelimella.

Historian laitos ja yleisen kielitieteen laitos julkaisevat HTML-versiota laitoksen tiedotuslehdestä.

#### **2.1.4 Matemaattis-luonnontieteellinen tiedekunta**

Geofysiikan laitoksen elektronisesta julkaisusarjasta *WWW Notes on Geophysics* on ilmestynyt HTML-muodossa yksi numero marraskuussa 1997. Se on tehty Netscapen Composer-editorilla.

Maantieteen laitoksen pro gradu -tutkielmista neljä on julkaistu HTML-muodossa verkossa. Näistä kaksi jälkimmäistä löytyvät myös elektronisen julkaisupankin palvelimelta. Näistä kahdesta toisen on tekijä tehnyt FrontPage-ohjelmalla ja toinen on muunnettu PageMakerilla HTML:ksi. Lisäksi laitoksen sivuilla on neljä artikkelia HTML-muodossa. Laitos julkaisee lisäksi verkossa tiedotuslehteä *WEB.DOC*, josta on ilmestynyt 12 numeroa.

Matematiikan laitoksen julkaisemasta raporttisarjasta on verkossa 24 numeron Postscript-versiot. Julkaisut on tehty LaTeX:lla. Lisäksi julkaisusarjasta *Annales Academiæ Scientiarum Fennicæ* on ilmestynyt 6 numeroa elektronisessa muodossa sekä postscript että (La)TeX:n dvi-tiedostoina.

Meteorologian laitoksen julkaisemsta raporttisarjasta on verkossa ilmestynyt kaksi numeroa Postscript-tiedostoina. Nämä on tehty LaTeX:lla.

Tietojenkäsittelytieteen laitoksen julkaisemasta raporttisarjasta on verkossa saatavilla 99 numeroa Postscript-tiedostoina. Nämä on tehty LaTeX:lla. Lisäksi yksi väitöskirja on julkaistu PDF-muodossa. Tämä on tehty ilmeisesti LaTeX:lla.

#### **2.1.5 Kasvatustieteellinen tiedekunta**

Joistakin Kasvatustieteen laitoksen julkaisuista on saatavilla HTML-muotoinen tiivistelmä.

#### **2.1.6 Valtiotieteellinen tiedekunta**

Käytännöllisen filosofian laitos on julkaissut oppaan graduntekijöille HTML-muodossa.

## **2.2 Muita vastaavia koti- ja ulkomaisia hankkeita**

### **2.2.1 HUTPubl**

HUTPubl on teknillisen korkeakoulun rakenteinen julkaisutuotanto. Projektin tavoitteena on saada aikaan systeemi- ja palvelukokonaisuus Teknillisen korkeakoulun julkaisusarjojen rakenteiseen tuottamiseen, elektroniseen arkistointiin ja verkkojakeluun sekä paperimuotoisen julkaisun tuottamiseen tarvittaessa. Projektisuunnitelma on nelivuotinen. Projekti on saanut rahoitusta Opetusministeriön Suomi tietoyhteiskunnaksi -ohjelmasta. 1. jatkoprojekti alkoi 21.3.1997 ja päättyi 30.4.1998. Kotisivu on osoitteessa: <http://www.hut.fi/Yksikot/Kirjasto/HUTpubl/> ja se sisältää mm. projektin SGML-tyylimäärytykset ja testituloksia. Testauksessa kaksi koehenkilöä kokeili rakenteisen dokumentin tuottamista FrameMakerin SGML-editorilla ja yksi testaajista kokeili Wordia. Testaajien atk-kokemus vaihteli. Lopputuloksena oli, että ns. WYSIWYG ohjelmilla kuten Word on rakenteisen dokumentin tekeminen vaivalloisempaa kuin erityisesti tätä varten tehdyllä ohjelmalla.

### **2.2.2 Oulun yliopiston kirjasto – elektroniset julkaisut**

Oulun yliopiston *Acta*-sarjan väitöskirjat julkaistaan elektronisessa muodossa HTML- ja PDF-dokumentteina. Palvelun kotisivut ovat osoitteessa: <http://herkules.oulu.fi/>

### **2.2.3 Jyväskylän yliopiston kirjaston tutkielmapankki**

Jyväskylän yliopiston gradut 1.1.1997 lähtien sekä muutama muu julkaisu on talletettu JPEG-kuvina. Järjestelmänä on Xeroxin DigiPath. Palvelu löytyy osoitteesta: <http://docuweb.jyu.fi/>

### **2.2.4 The Nordic Centre of Excellence for Electronic Publishing (NordeP)**

Elektronisen julkaisemisen pohjoismainen keskus, joka selvittää elektronisen julkaisemisen tilaa pohjoismaissa ja pyrkii antamaan suosituksia. Alullepanijana on NordINFO ja toteuttajana VTT. Kotisivu on <http://www.vtt.fi/inf/nordep/index.htm> Sisältää mm. linkin UNESCO:n ohjeisiin kirjastojen WWW-sivujen tuottamiseksi. Lisäksi linkkikokoelma, jonka aiheena on Adobe Acrobat ja PDF-tiedostoformaatti.

Keskuksen toimintaan kuuluu kolme projektia: WEBPILOT (tutkii verkkojulkaisemista), NORD-CD (tutkii CD-ROM –julkaisemista) ja OCR-Index (tutkii kirjastojen mikrofilmattujen sanomalehtien muuttamista tekstitiedostoiksi). Keskus on julkaissut selvityksiä HTML- ja PDF-muodoissa.

### **2.2.5 Linköping University Electronic Press**

1996 aloitettu Linköpingin yliopiston elektroninen julkaisupalvelu. Julkaisut talletetaan Postscript-, PDF- tai HTML-formaatissa. Toistaiseksi julkaisut ovat lähinnä tiedeartikkeleita ja elektronisia lehtiä. Opinnäytetöitä ei vielä ole julkaistu. Artikkelit kootaan julkaisupalvelun ylläpitämiin elektronisiin sarjoihin, joilla kullakin on oma toimittajansa.

Kotisivu on osoitteessa: <http://www.ep.liu.se/> ja sieltä löytyvät mm. ohjeet ja valmiit tyylimääritykset kirjoittajille sekä sopimus, joka kirjoittajien täytyy allekirjoittaa.

### **2.2.6 The Internet Project**

Göteborgin yliopiston Internet-projekti. Kotisivu on osoitteessa: <http://internet.adb.gu.se/> Projekti on julkaissut joukon artikkeleita, jossa käsitellään myös verkkojulkaisemisen ongelmia. Artikkelit ovat HTML-, Postscript- tai PDF-muodossa. Projekti jakaantuu eri toimintoihin, joista *DocFlex* keskittyy aineiston välittämiseen verkon kautta.

### **2.2.7 NYST-Informasjonsformidling**

Helmikuussa 1997 päättynyt Oslon yliopiston informaatioteknologiakeskuksen projekti, joka selvitti verkkojulkaisemiseen liittyviä asioita. Kotisivu on osoitteessa: <http://www.uio.no/usit/Utviklingsseksjonen/nyst/nyst.html> ja sieltä löytyy joukko norjankielisiä selvityksiä ja raportteja.

SGML-core on Oslon yliopiston SGML-projekti. Kotisivu osoitteessa: <http://www.uio.no/usit/ADB-seksjonen/sgml-core/>

## 2.2.8 NDLTD — Networked Digital Library of Theses and Dissertations

26 amerikkalaisen yliopiston yhteinen elektroninen opinnäytetyöpankki. Projektia koordinoi Virginian teknillinen korkeakoulu. Kotisivu on osoitteessa: <http://www.ndltd.org/>

Virginian teknillinen korkeakoulu on toimittanut ohjeita kirjoittajille. Ks. <http://etd.vt.edu/>

## 2.2.9 UW Electronic Theses and Dissertations Server

Waterloon yliopiston (Kanada) elektroninen opinnäytetyöpankki. Julkaisee ainoastaan PDF-muodossa. Osa töistä on skannattu paperikopioista. Kotisivu on osoitteessa: <http://www.lib.uwaterloo.ca/~uw-etpt/pilot.html>

## 2.3 Väitöskirjan julkaiseminen

Julkaisupankkihankkeen yhtenä keskeisenä tavoitteena on kokeilla väitöskirjojen elektronista julkaisemista painetun väitöskirjan rinnalla. Kokonaan elektroniseen väitöskirjaan siirtyminen voi tapahtua vasta nykyisen vapaakappalelain uudistamisen jälkeen, jolloin elektroniset julkaisut sisällytetään vapaakappaleina luovutettaviin ja pitkäaikaisesti säilytettäviin aineistoihin. Vapaakappalelain uudistusta valmisteleva opetusministeriön asettama työryhmä sai työnsä päätökseen kesäkuussa 1998 ja lain valmistelu jatkuu vuoden 1999 eduskuntavaalien jälkeen.

Väitöskirjojen elektronista julkaisemista on selvitetty erillisessä työryhmässä, jonka tehtävänä on ollut laatia malli väitöskirjojen elektronista julkaisuprosessia varten. Tätä varten työryhmä on selvittänyt mm. perinteisen julkaisuvaihdon korvaamista tiedotuksella elektronisista julkaisuista ja elektronisen väitöskirjan julkaisuohjeita ja tyylimäärityksiä.

Väitöskirjojen nykytilaa ohjaavia säädöksiä ovat:

- Helsingin yliopiston opintasuoritusten arvostelua ja kuulusteluja sekä oikeusturvalautakuntaa koskeva johtosääntö, kanslerin vahvistama 30.4.1992
- Yliopistolle luovutettavien väitöskirjojen määrää ja jakelua pohtineen työryhmän muistio 18.5.1992. Aiheeseen liittyvä konsistorin päätös 14.10.1992.
- Kirjastotoimikunnan esitys konsistorille 9.2.1978 tietojen toimittamisesta Helsingin yliopiston hyväksymistä väitöskirjoista kansainväliseen väitöskirjojen tiivistelmäjulkaisuun, *Dissertation Abstracts International*, Series C. Konsistorin asiaan liittyvä päätös.

Oikeusturvajohtosäännön mukaan väitöskirjan tulee olla painettu tai muulla vastaavalla tavalla valmistettu. Tämä sanamuoto ei mahdollista pelkästään elektronisessa muodossa julkaistavaa väitöskirjaa. Johtosäännön uudistus on käynnissä, ja



julkaisupankki -hanke on valmistellut elektronisen julkaisemisen mahdollistavaa kannanottoa johtosäännön muutokseen.

Luovutettavien väitöskirjojen nykyinen määrä on 160 kappaletta ja artikkeliväitöskirjoilla 30 kappaletta. Väitöskirjoista 100 (10) kappaletta luovutetaan yliopiston kirjastolle, ja 60 (20) kappaletta yliopistolle.

Yliopiston kirjasto sisällyttää omiin kokoelmiinsa 3 kappaletta kutakin väitöskirjaa ja jakaa muille vapaakappalekirjastoille 5 kappaletta. Loput toimitetaan julkaisuvaihtoina koti- ja ulkomaisille kirjastoille. Vuosittain jaeltavien vaihtokappaleiden määrä on noin 15 000 julkaisua. Vaihtokumppaneina on 158 organisaatiota, joista 32 on kotimaisia. Kirjastolle julkaisuvaihtojen korvaaminen tiedotuksella elektronisista väitöskirjoista tuottaisi säästöjä henkilötöyssä ja aineiston varastoinnissa. Tieto uusista väitöskirjoista saataisiin kuitenkin leviämään aikaisempaa laajemmalle ja dokumentit olisivat käytettävissä nykyistä nopeammin. Vaihtojen lakkauttamisen vaikutuksia omien kokoelmien kartuntaan ja verkkopalvelujen ulottumattomilla olevien vaihtokumppanien tiedonsaantiin ei kuitenkaan ole pystytty tarkasti arvioimaan.

Väitöskirjatyöryhmä on tehnyt tiedekunnille kyselyn, jossa on pyydetty kannanottoa painettujen väitöskirjojen määrän vähentämiseen, mikäli väitöskirjasta julkaistaisiin myös elektroninen versio. Väitöskirjojen kokonaismääräksi ehdotetaan artikkeliväitöskirjojen nykyisen käytännön mukaista 30 kappaletta, josta 6 menisi kirjastolle, 4 keskushallinnolle ja 20 kappaletta tiedekunnalle. Vuoden vaihteeseen mennessä saatiin vastaukset viideltä tiedekunnalta, joista lääke- ja oikeustieteellinen sekä humanistinen tiedekunta olivat valmiita määrän alentamiseen ehdotuksen mukaisesti, maatalous-metsätieteellinen tiedekunta on valmis vähentämään määrän 24 kappaleeseen, teologinen ja matemaattis-luonnontieteellinen tiedekunta halusivat säilyttää nykyisen määrän ja valtiotieteellinen tiedekunta ilmoitti uudistavansa julkaisukäytäntöään ja vastaavansa myöhemmin. Kasvatustieteellisen ja eläinlääketieteellisen tiedekunnilta on saatu vuoden 1999 puolella kannanotot, joissa periaatteessa ei vastusteta väitöskirjojen julkaisemista elektronisesti, mutta vähimmäismäärän supistamiseen ei varsinaisesti otettu kantaa. Keskustelua tiedekuntien kanssa jatketaan, jotta saataisiin tarvittava yksimielisyys konsistorille tehtävää, painettujen rinnakkaisjulkaisujen määrän alentamista suosittelevaa ehdotusta varten.

### 3. Sopimukset

Elektronisessa julkaisupankissa julkaistavat teokset ovat tekijänoikeuden alaisia ja niiden jakeleminen verkossa edellyttää kirjallisia julkaisusopimuksia. Sopimusten laadinnassa asiantuntijoina ovat toimineet apulaisprofessori Antti Kivivuori ja yliassistentti Rainer Oesch. Asianajotoimisto Borenus & Kempiseltä on hankittu projektin käyttöön Samuli Simojoen opiskelijakirjaston Opintoaineistot verkossa -hanketta varten laatima raportti *Oppimateriaalin verkkojulkaisemisen oikeudelliset kysymykset*.

Laaditun julkaisusopimuksen tarkoituksena on määritellä teoksen tekijän ja yliopiston kirjaston oikeudet ja velvollisuudet dokumentin siirtoa, elektroniseen muotoon saattamista, verkon välityksellä yleisön saataville asettamista ja arkistointia varten. Korvausten osalta sovitaan, että kirjasto ei maksa tekijälle korvausta teoksen käytöstä eikä tekijä maksa palkkiota kirjaston suorittamasta työstä. Julkaisupankin ja sen sisältämien teosten käyttö yleisölle on maksutonta.

Sopimuksen mukaan teoksen tekijänoikeudet kuuluvat tekijälle. Kirjastolla on oikeus muuntaa teos palvelun edellyttämiin tiedostomuotoihin ja tarjota se yleisön käyttöön sopimuksen voimassaolon ajan. Mikäli tekijä irtisanoo sopimuksen, jää kirjastolle oikeus säilyttää kokoelmissaan julkaisun arkistokappale.

Tiedot julkaisupankkiin sisältyvien julkaisujen tekijänoikeuksista tallennetaan dokumenttien kuvailutietoihin.

## 4. Julkaisuympäristö

Hankkeen käytössä on 200 Mhz:n Pentium II mikrotietokone, jossa on 64 Mt muistia sekä 4 Gt kovalevytilaa. Kovalevy on jaettu kolmeen partitioon, 2 Gt on varattu ensisijaisesti käyttöjärjestelmälle ja sovellusohjelmille, 1 Gt on varattu hankkeessa käsiteltävälle aineistolle ja dokumentaatiolle. Lisäksi käytössä on 1 Gt ylimääräistä varastointitilaa. Tulostustarpeita varten on käytössä HP LaserJet 6P. Tällä voidaan testata mm. PDF-tiedostojen tulostumista paperille. Tarvittaessa käytössä on kirjastotoimen verkkopalvelujen kehittämisprojektien PowerPC, jossa käyttöjärjestelmänä on MacOS. Macintoshissa on käytössä mm. PageMaker 6.5.

Käyttöjärjestelmänä on Windows 95. Lisäksi koneessa on dokumenttien käsittelyä ajatellen on seuraavat sovellukset: Microsoft Office 97 Professional, Corel WordPerfect Suite 8 ja MikTeX. Selaimina on sekä Netscape Communicator 4.5 että Microsoft Internet Explorer 5.0:n beta-versio. HTML-editorina on Arachnophilia. Kuvankäsittelyssä on käytetty ohjelmia Paint Shop Pro 4.0 ja Corel PhotoPaint 7.0. Postscript-tiedostoja on käsitelty Ghostscriptillä ja Ghostviewillä. Näillä on myös muunnettu EPS-kuvia bittikarttakuviksi tarvittaessa. Tekstieditorina on käytetty GNU Emacsin Windows-versiota.

PDF-tiedostojen valmistamista kokeiltiin ensin ilmaisella Ghostscriptillä, mutta myöhemmin siirryttiin Adobe Acrobat ohjelmistoon.

LaTeX-dokumentit kääntämiseksi HTML:ksi kokeiltiin eri sovelluksia. TTH sisältyy MikTeX-järjestelmään. Se toimii nopeasti, mutta ei pysty kääntämään yliopistotasoisia matematiikkaa. LaTeX2HTML on UNIX-ympäristöissä suosittu järjestelmä, joka koostuu Perl-skripteistä. Siitä kokeiltiin useita versioita ennen kuin se saatiin toimimaan Windows 95 -järjestelmässä. Windows-versiot tuntuvat toimivan ainoastaan Windows NT -ympäristössä. DOS-versio toimi, mutta matematiikan pro gradu oli sille liian vaikea ja ohjelma kaatui kesken työn. Myös Atk-osaston UNIX-palvelimissa olevaa versiota kokeiltiin, mutta ohjelmaa ei saatu

toimimaan saman, mutkikkaan, pro gradun kanssa. Atk-osasto ei vastaa ohjelman ylläpidosta. LaTeX2HTML:n ongelmana on lisäksi se, että se ei tunnista käyttäjän määrittelemiä makroja eikä eksoottisempia makropakkauksia. Näistä pitäisi kai kista tehdä Perl-kieliset versiot. Lopulta DVI4TeX onnistui kääntämään työn. Tämä ohjelma tekee käännöksen DVI-tiedostosta eikä .tex-tiedostosta. Niinpä työssä voidaan käyttää myös omia makroja. Lopputulos käyttää CSS-tyylejä. Yksinkertaiset yhtälöt esitetään HTML-muodossa ja hankalammat GIF-kuvina. Lopputulos vaatii kuitenkin käsin tehtävää korjailua ja oikolukua, joka vie aikaa useita työpäiviä tehokasta työaikaa.

## **5. Julkaisuohteet ja julkaisutuotanto**

### **5.1 Pilottiaineisto**

Julkaisupankin WWW-sivuilla on yleisohjeita siitä, mitä aineistoa julkaisupankin pilottihankkeeseen otetaan mukaan. Osoite on <http://hul.helsinki.fi/elbanco/julkaisut/index.html> Pilotissa testataan väitöskirjoja, liseniaattitöitä sekä syventävien opintojen tutkielmia. Opetusmateriaali jää pois, sillä Opiskelijakirjasto osallistuu hankkeeseen, joka selvittää näiden julkaisemista.

Kaikkea pilottiaineistoa ei välttämättä tulla julkaisemaan. Niitä kuitenkin käytetään testaamaan julkaisu- ja konversio-ohjelmia.

### **5.2 Yhtenäinen ulkoasu**

Eräs Julkaisupankin tavoitteista on luoda elektronisille julkaisuille yhtenäinen ulkoasu eli yliopiston elektroninen julkaisusarja. Tätä varten Oy Edita Ab:sta tul- laan vuoden 1999 aikana hankkimaan ostopalveluna tarvittavat ulkoasun määri- tykset sekä valmiit tyyliohjat tuettaville ohjelmistoille.

### **5.3 Tuettavat ohjelmistot**

Elektroninen julkaisupankki tukee ohjelmia, jotka keväällä 1998 tehdyssä kyse- lyssä todettiin yleisimmin käytetyiksi Helsingin yliopistossa. Tekstinkäsittelyoh- jelmista tuetaan Microsoft Wordia versioita 6.0-8.0, WordPerfectistä versioita 5.0- 8.0. Taitto- ja ladontaohjelmista tuetaan ensisijaisesti LaTeXia sekä jossain mää- rin PageMakeria.

Yllä mainitut tiedot löytyvät WWW-sivuilta osoitteesta:  
<http://hul.helsinki.fi/elbanco/ohjeet.html> lisäksi osoitteessa  
<http://hul.helsinki.fi/elbanco/infobank.html> on lisätietoa erilaisista julkaisu- ja konversio-ohjelmista, jotka soveltuvat elektroniseen julkaisemiseen.

## 5.4 DC-kuvailut ja URN-generaattori

Elektronisen julkaisupankin HTML-dokumentteihin liitetään Dublin Core -määritysten mukaiset metadatakentät. Näin luettelointitiedot voidaan sisällyttää itse dokumenttiin ja tiedot voidaan tarvittaessa poimia ohjelmallisesti viitetietokantoihin. Lisäksi dokumenttien löytyvyys paranee, kun verkkoon tulee hakukoneita, jotka osaavat hyödyntää metadataa.

Dokumenteille annetaan myös URN-tunnus. Tämä on Internetin vastine paperisen maailman ISBN-numeroille eli se identifioi yksiselitteisesti dokumentin verkossa, vaikka dokumentin osoite eli URL muuttuisikin. Tunnuksen hyödyntäminen tosin edellyttää, että verkossa on käytössä hakukoneita, jotka osaavat hakea dokumentin URN-tunnuksen avulla.

URN-tunnus määritellään erityisen URN-generaattorin avulla. Tämä on verkossa oleva CGI-sovellus joka jakaa URN-tunnuksia niin, ettei samaa tunnusta anneta kahdelle eri dokumentille. Suomen tunnuksia jaetaan osoitteesta:  
<http://linnea.helsinki.fi/cgi-bin/urn.pl>

## 5.5 Ohjeet kirjoittajille

Toistaiseksi kirjoittajille on ollut tarjolla muutamia vain aivan yleisiä ohjeita osoitteesta: <http://hul.helsinki.fi/elbanco/ohjeet/wordohje.htm> Ohjeet on suunnattu Wordin käyttäjille, mutta samat asiat koskevat myös WordPerfectin käyttäjiä. Ohjeissa on mainittu asioita, joita kirjoittajan tulisi välttää silloin, kun työstä tehdään myös HTML-versio:

- dokumenttia ei tule tavuttaa käsin, sillä HTML-dokumenteissa sivun koko ei ole vakio
- taulukot tulee tehdä ohjelman taulukko-toiminnolla eikä sarkainnäppäimellä tai välilyönneillä (tekstinkäsittelyohjelma ei ole kirjoituskone), jotta ne voidaan konvertoida HTML-taulukoiksi
- sisennykset tulee tehdä tyyliellä, eikä sarkain- tai välilyöntinäppäimillä. Sarkainnäppäimen käyttöä tulee välttää, sillä sarkaimilla tehty taitto ei säily HTML-konversiossa.

Varsinaiset laajemmat ohjeet on tarkoitus tehdä sitten, kun ostopalveluna hankittavat taittomääritykset ovat valmiit eli vuoden 1999 puolella.

## 5.6 Tiedekuntien toimittama aineisto

Vuoden 1998 aikana julkaisupankki sai tiedekunnilta 8 pro gradu -tutkielmaa sekä kolme väitöskirjaa. Yksi syventävien opintojen tutkielma oli eläinlääketieteellisestä tiedekunnasta, loput olivat matemaattis-luonnontieteellisestä. Väitöskirjat olivat lääketieteellisestä, humanistisesta sekä matemaattis-luonnontieteellisestä tiedekunnasta, yksi kustakin.

Toimitetuista töistä kolme oli tehty LaTeXilla, kolme oli tehty Wordilla, yksi WordPerfectillä ja yksi PageMakerilla (Macintosh-versio). Yksi työ toimitettiin valmiina HTML-dokumenttina ja kaksi valmiina PDF-tiedostona.

Aineisto vaati HTML-konversion jälkeen käsin tehtävää korjailua. Varsinkin kuvissa esiintyvä teksti tulee helposti liian suttuisena, jolloin se on tehtävä kuvankäsittelyohjelmassa uudelleen. Jos alkuperäiset kuvatiedostot ovat saatavilla, voi kuvat yrittää konvertoida grafiikkaohjelmilla HTML-dokumenteissa käytettävään muotoon. Wordin HTML-konversio näyttää toimivan parhaiten kun alkuperäinen työ on tehty Word97:lla ja työssä on käytetty Wordin tyylimääryityksiä. WordPerfectistä ei saatu vuoden 1998 aikana vielä tarpeeksi kokemuksia, jotta sen dokumenttien konvertointimahdollisuuksia voitaisiin kommentoida. Saamamme LaTeX-dokumentit olivat matematiikan syventävien opintojen oppinnäytteitä, jotka sisältävät huomattavissa määrin matemaattisia yhtälöitä. Näiden konvertoinnissa ohjelmat olivat ilmeisesti liiankin lujilla, koska konversion jälkeen koko teksti on vielä käytävä läpi käsin ja korjattava virheet HTML-editorilla. Matematiikan konvertointi muistuttaa jossain määrin OCR-skannausta. Siinäkin tekstintunnistusohjelman virheprosentti on yleensä niin suuri, että teksti on vielä oikoluettava.

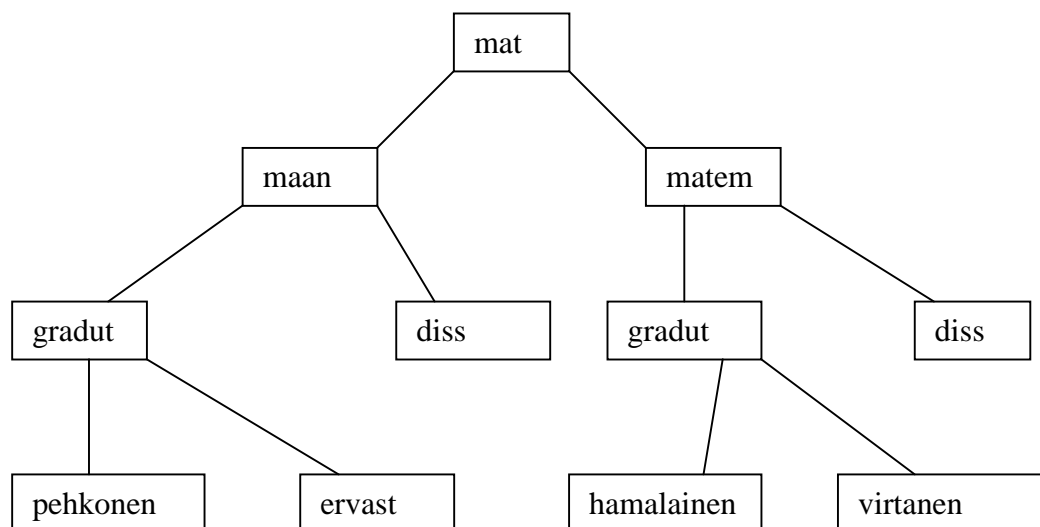
Tarkempia tietoja konvertointiin liittyvistä ongelmista on saatavissa dokumentista *HY:n elektronisen julkaisupankin saamat dokumentit*, joka löytyy verkosta osoitteesta: <http://hul.helsinki.fi/elbanco/aineisto.html> tai <http://hul.helsinki.fi/elbanco/aineisto.pdf>

Yhteenvetona todettakoon, että nykyisten tekstinkäsittelyohjelmien HTML-konversiossa on ongelmia, jotka eivät mahdollista niiden käyttöä täysin automatisoidussa ympäristössä. Lopputulos joudutaan aina oikolukemaan ja korjaamaan käsin. Vuoden 1999 aikana Microsoft julkaisee Wordista uuden version, jossa HTML-konversion pitäisi olla kehittyneempi. Ennakkotietojen pohjalta vaikuttaisi siltä, että ainakin Wordin konversio-ongelmat olisivat tämän jälkeen historiaa. Taitto-ohjelmista kokeiltiin PageMakeria, jonka HTML-konversiossa ei ollut valittamista. Se noudatti PageMakerin tyylejä ja käyttäjä pystyi itse määrittelemään miten konversio tehtiin. LaTeX-dokumenteissa ongelmana on se, että ne sisältävät yleensä hyvin paljon matematiikkaa, jonka esittäminen HTML-ympäristössä on ongelmallista. Niinpä ongelmat eivät johdu pelkästään konversio-ohjelmista, vaan myös HTML-kielen rajoittuneisuudesta. Konversio-ohjelmat olivat lisäksi hankalia käyttää.

## 6. Julkaisujen arkistointi

Saatu aineisto on tallennettu Helsingin yliopiston kirjaston UNIX-palvelimelle. Aineistosta pyritään tallentamaan HTML- ja PDF-versioiden lisäksi myös alkuperäiset tiedostot, joista konversiot on tehty. Lisäksi mukaan tallennetaan lyhyt tekstimuotoinen tiedosto, joka sisältää aineiston mukana toimitettavan saatelomakkeen tiedot eli tiedot kirjoittajan käyttämistä ohjelmista ja tiedostomuodoista.

Aineisto on jaettu palvelimella eri alihakemistoihin siten, että ylimmällä tasolla on tiedekunta, tämän alapuolella on laitos ja laitoksen sisällä tiedostot jaetaan vielä väitöksiin, lissensiaattitöihin sekä pro gradu -tutkielmiin. Näiden alimpien hakemistojen sisällä aineisto jaetaan tekijän sukunimen mukaisiin hakemistoihin.



*Yllä on esimerkki hakemistorakenteesta.*

Aineiston siirto tapahtuu niin, että yhdyshenkilö tai kirjoittaja postittaa työnsä levykkeellä sekä paperille tulostettuna julkaisupankille. Pienehköjä tiedostoja on postitettu myös sähköpostitse. Lisäksi töitä on toimitettu niin, että kirjoittaja on pannut sovittuna aikana työnsä kopiot mikroverkon temp-hakemistoon, josta työ on samana päivänä noudettu julkaisupankkiin.

Tulevaisuudessa tiedostojen siirto kannattaisi tehdä mahdolliseksi myös linjasiirtona jollekin julkaisupankin palvelimelle. Yliopistopaino käyttää FTP-palvelinta tähän tarkoitukseen. FTP:n tietoturvasa on nykyisellään kuitenkin parantamisen varaa ja tästä syystä on selvittävää myös muita tapoja tiedostojen siirtämiselle.

## 7. Koulutus ja tiedotus

Elektroninen julkaisupankki järjesti 27.10.1998 koulutustilaisuuden tiedekuntien yhdyshenkilöille ja muille toimintaan liittyville henkilöille. Koulutuspäivä koostui

kahdesta osasta. Aamupäivällä osallistuimme VTT Tietopalvelujen kursille *Verkkojulkaiseminen* ja iltapäivällä suunnittelija Timo Hellgren piti yhdyshenkilöille tiedotustilaisuuden.

Aamupäivän kursilla käsiteltiin seuraavia aiheita:

- julkaisuprosessin vaiheet
- tiedostoformaatit: HTML, Postscript, PDF ja SGML/XML
- HTML-dokumenttien suunnittelu ja kirjoittaminen
- Acrobat-ohjelmat ja PDF-dokumenttien tuottaminen

## 8. Määrärahan käyttö

Hankeen kokonaisbudjetti vuodelle 1998 oli 400 000 markka, josta yliopiston osuus oli 300 000 markkaa ja kirjaston omavastuuosuus 100 000 markkaa. Eritelty käyttösunnitelma tehtiin yliopiston myöntämän määrärahan osalta. Kirjaston oma osuus sisältää henkilötyötä.

Yliopiston myöntämästä määrarahasta on käytetty yhteensä 158 063,46 markkaa, joka on 53% suunnitellusta. Henkilötyön osuudesta on toteutunut 69%, mutta laitteiden ja ohjelmistojen sekä palvelujen ostoihin on käytetty vain neljännes suunnitellusta määrarahasta. Myös muut menot ovat olleet ennakoitua pienemmät.

Määrärahan käyttö vuonna 1998

Menolaji	Suunnitelma	Toteutuma	%
Henkilötyö	190 000	131 595,39	69%
Laitteet, ohjelmistot	60 000	14 954,92	25%
Koulutus, palvelut	40 000	10 891,28	27%
Muut menot	10 000	622,52	6%
Yhteensä	300 000	158 063,46	53%

Määrärahojen käytön alitus johtuu hankeen edellyttämästä ennakoitua suuremmasta taustaselvitysten tarpeesta. Myös julkaisukäytäntöihin ja normeihin liittyvät päätökset ovat olleet edellytyksenä, jotta pilottia voidaan laajentaa. Kun julkaisusopimukset valmistuvat ja saadaan allekirjoitettua, voidaan julkaisujen jakelu verkossa käynnistää. Tämä todennäköisesti parantaa hankkeen näkyvyyttä ja edesauttaa testiaineiston saatavuutta, jolloin kokeilua voidaan laajentaa. Julkaisujen käsittely ja tuottaminen elektroniseen jakelumuotoon ovat työvoimaintensiivisiä prosesseja ja edellyttävät nykyistä useampia henkilöitä ja laitteita projektin käyttöön. Lisätyövoiman rekrytointi ja laiteinvestoinnit on tästä syystä siirretty toiselle toimintavuodelle, jolloin varsinainen julkaisutuotanto saadaan käyntiin. Vuodelle 1999 siirtyvien määrärahojen osuus on 141 935,54 markkaa.

Kirjaston oma henkilötöyöosuus hankkeessa on ollut projektin johtaja 1 kuukausi, projektisihteeri 2,5 kuukautta ja sovellussuunnittelijan asiantuntijatyötä 3 kuukautta. Hankerahoituksella palkattuna on ollut kokopäivätoiminen suunnittelija.